



A Feature Based Approach to Leveraging Context for Classifying Newsgroup Style Discussion Segments

Yi-Chia Wang, Mahesh Joshi, Carolyn Rosé

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Motivation

Information overload in communication media (email, instant messaging, discussion boards)

↓ Solution

Conversation summarization

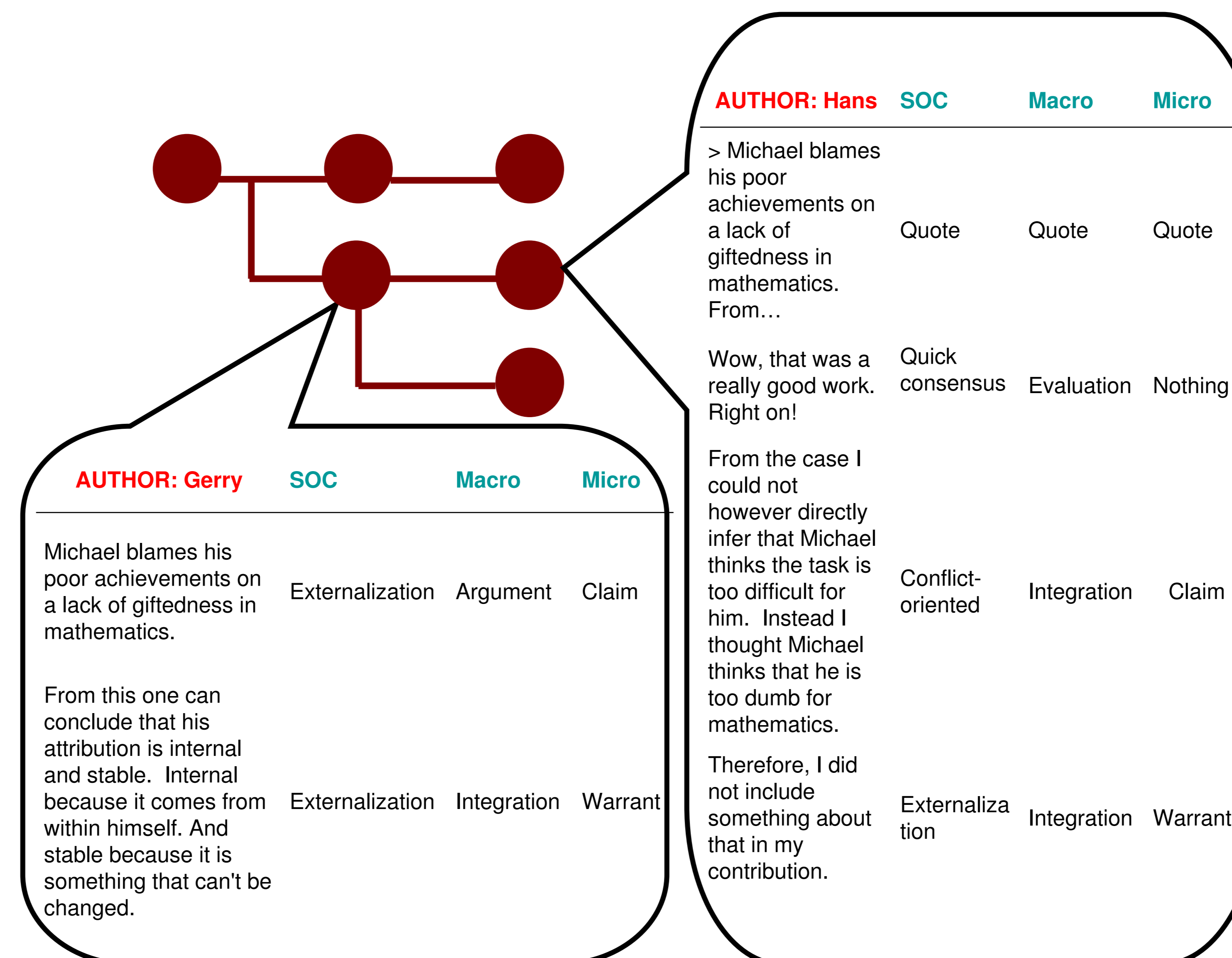
Content + Style + Structure

↓ Intermediate Step

Classifying segments of conversation

Goal

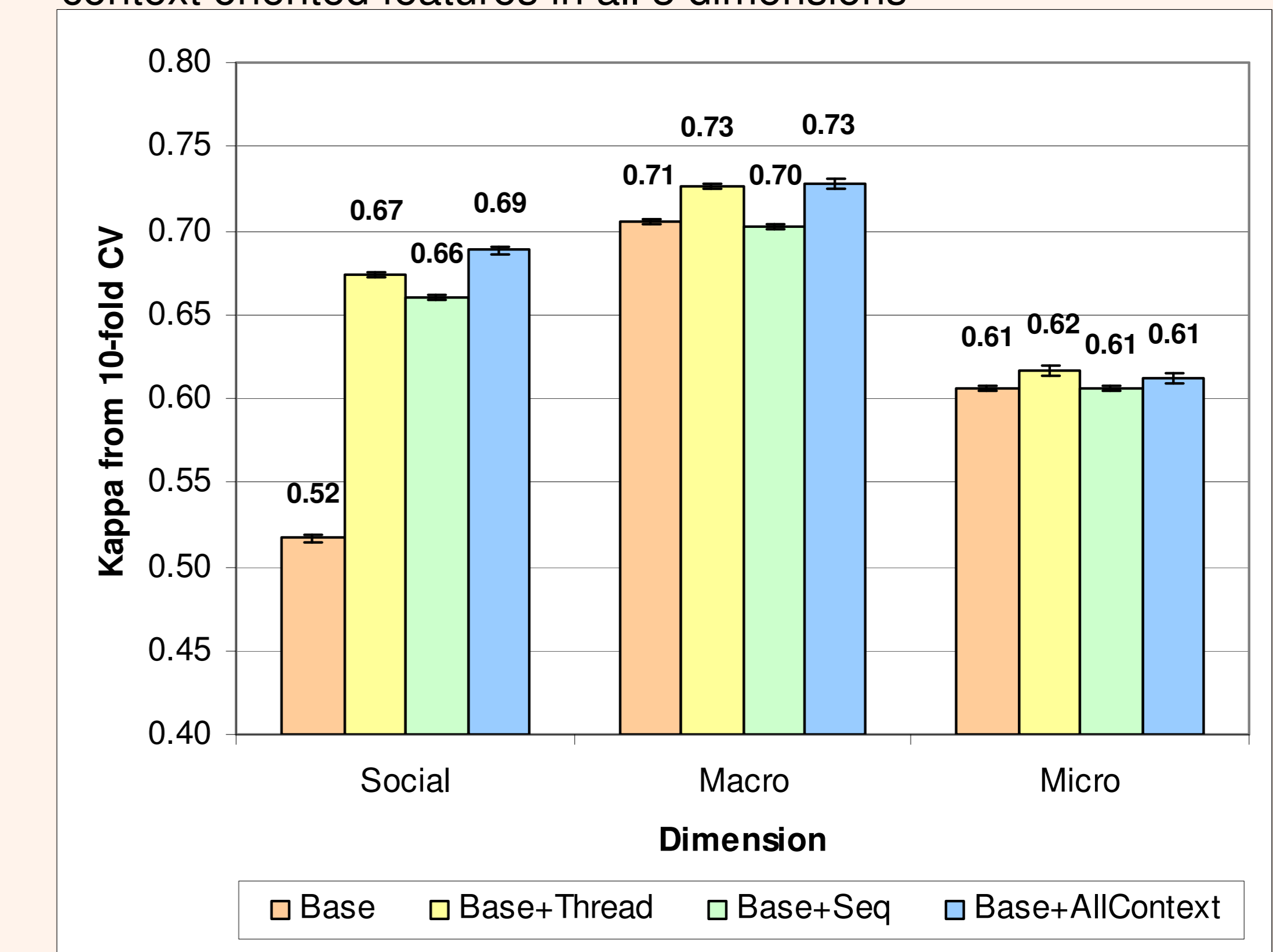
- Assessing the quality of newsgroup style interactions using a multi-dimensional annotation scheme
- Leveraging the structure and style of discussion forums
 - Novel thread based features
 - Sequential data



Results

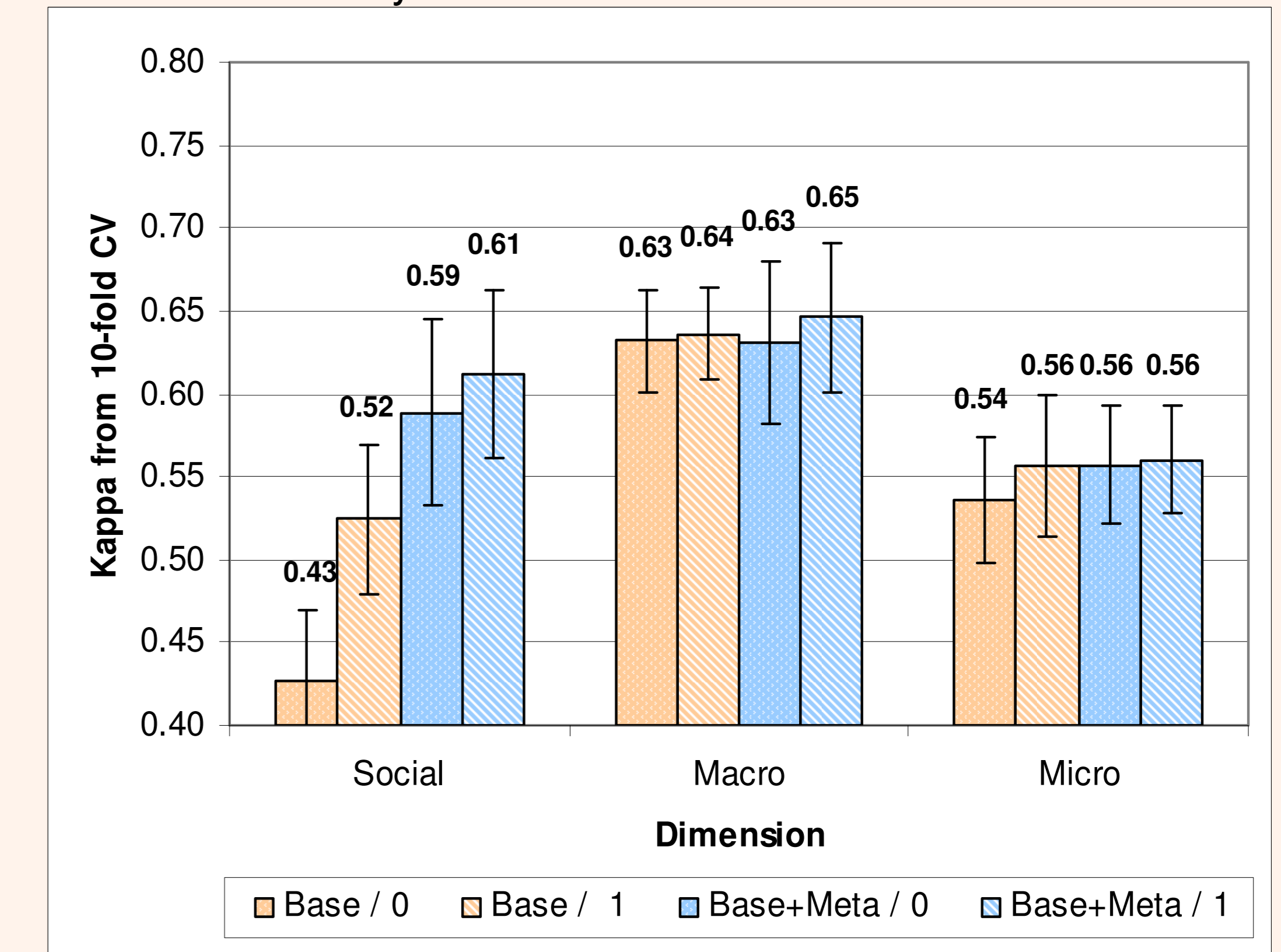
Alternative Feature Sets (SVM)

Achieving a statistically significant improvement by adding context oriented features in all 3 dimensions



Sequential Learning (Collins Perceptron)

Sequential learning only has a significant effect on one dimension, and even there only in the absence of context oriented features.



Note: Cohen's Kappa Statistic is a commonly used metric in assessing inter-rater reliability on the defined analyses units among target categories.

Features

Baseline

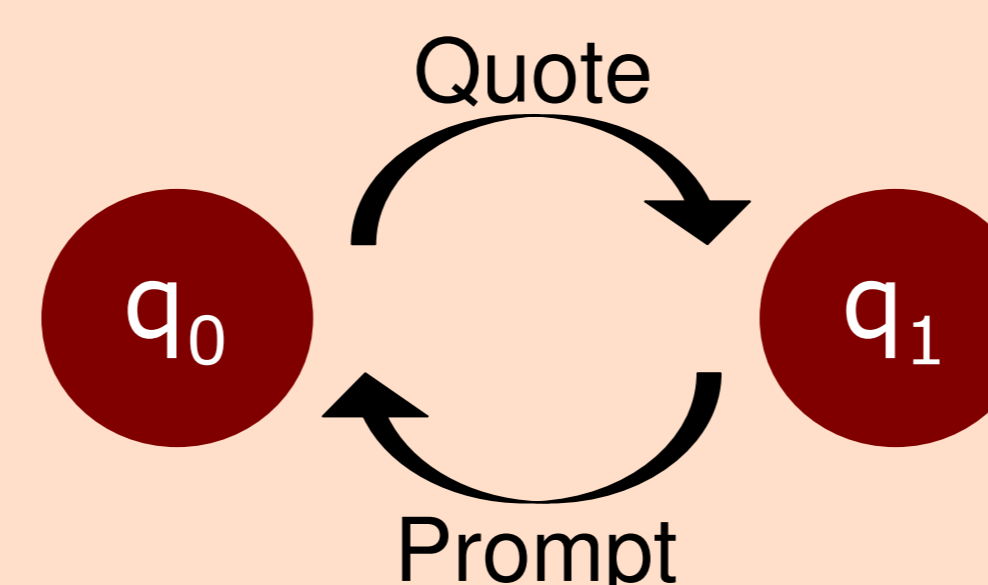
- Un-stemmed unigrams
- Punctuation
- Removing features that occurred less than 5
- Length of each segment

Thread Structure Features

- Deep: the depth in the thread where a message appears
- Parent and Child: semantic similarity between current message to the parent message

Sequence-Oriented Features

Constructing a finite-state automaton having two states



Data and Annotation Scheme

Our data is from course discussion forums where students are discussing case studies. Messages are segmented into **idea units** and coded with 3 context-oriented dimensions.

Dimension	Categories
<ul style="list-style-type: none"> Micro-level of argumentation: Assessing the quality of individual arguments 	Simple claim, Qualified claim, Grounded claim, Grounded & qualified claim
<ul style="list-style-type: none"> Macro-level of argumentation: Examining the connection between individual arguments 	Argument, Counterargument, Integration/Reply, Question, Planning, Evaluation
<ul style="list-style-type: none"> Social Modes of Co-Construction: Referring to the degree of contributions of learners. 	Externalization, Elicitation, Quick consensus building, Integration-oriented consensus building, Conflict-oriented consensus building

Conclusions

Our work shows that novel thread based features have a consistently significant positive impact on the accuracy of classification of spans of conversational text. Thus, features such as these may serve as an important intermediate step towards conversation summarization.